

HTTP request and a receipt of a response to said HTTP request;

identifying a subset of said Internet resources that are candidates for prefetching based on said one or more estimated round trip times; and

determining whether to prefetch one or more of said Internet resources in said subset of Internet resources based on predefined conditions, at least one of said predefined conditions being based on a descending order said one or more estimated round trip times.

REMARKS

This Amendment is submitted in response to the outstanding Office Action, dated March 3, 2003 and is accompanied by a petition fee for extension of time (one month). Claims 1 through 29 are presently pending in the above-identified patent application. Claims 1, 3-7, 17, 19-22, 25, and 27-29 are proposed to be amended. No additional fee is due.

In the Office Action, the Examiner rejected claims 1 through 26 under 35 U.S.C. §103(a) as being unpatentable over Horvitz et al. (United States Patent Number 6,067,565) in view of Bryant et al. (United States Patent Number 6,078,956). The Examiner also rejected claims 1 through 29 under 35 U.S.C. §103(a) as being unpatentable over Kunkel et al. (United States Patent Number 5,961,603) in view of Narayanaswami (United States Patent Number 6,182,113) and Bryant et al. and rejected claims 1 through 29 under 35 U.S.C. §103(a) as being unpatentable over Kunkel et al. in view of Vaid et al. (United States Patent Number 6,119,235) and Bryant et al.

The present invention is directed to methods and apparatus for prefetching Internet resources based on the estimated round trip time for the resources referenced in a currently displayed Web page. Documents with the longest access times are prefetched first and prefetching generally continues until the estimated round trip time falls below a predefined-threshold. Thus, if a user clicks on an embedded hyperlink, the referenced document has either been fetched already or, if not, fetching the document from the origin server takes only a short time.

The "round trip" time or access time of a resource is the time interval between the sending of the first byte of an HTTP request for the resource until the last byte of the server response has arrived at the requesting Web client. The round trip time may be estimated in accordance with the present invention, for example, using an HTTP HEAD request. A HEAD request obtains status information and the size of the requested resource, *s*, from the origin server. If the server responds to the HEAD request with the document size, *s*, the prefetching agent computes the estimated round

trip time for further processing. If the server responds but fails to specify the document size, the prefetching agent utilizes the recorded average resource size, s , of resources previously received from the server. If the HEAD request does not yield any response from the server, or if the response shows an error code, then the prefetching agent determines that the hyperlinked document is not accessible and can provide an error message immediately once the user clicks on this hyperlink.

Independent Claims 1, 17, 25 and 27-29

Independent claims 1, 17, and 25 were rejected under 35 U.S.C. §103(a) as being unpatentable over Horvitz et al. in view of Bryant et al.

Although only claims 1 through 26 were rejected as being unpatentable over Horvitz et al. in view of Bryant et al., the Examiner asserts that, regarding claims 1, 25, and 27-29, Horvitz teaches prefetching Internet resources dependent on round-trip times based on send and receive times (col. 24, lines 12-20). Regarding claim 17, the Examiner asserts that Horvitz teaches “filtering web pages.” Col. 23, lines 19-26.

Applicant notes that Horvitz teaches that prefetching is performed according to an ascending order of prefetch times, i.e., an ascending order of the time required to prefetch an Internet resource. Horvitz teaches to “sort the page components in the set in descending order of their flux products. Once this occurs, block 1070 executes to fetch data from the URL component then exhibiting the highest flux product.” Col. 29, lines 39-42. Horvitz teaches that the flux product is equal to the product of the probability of the Internet resource being accessed, the value of the component and the transmission rate divided by the size of the component. See equation 14. Applicants note that the latter components of the flux product, i.e., the transmission rate divided by the size of the component, is equivalent to the inverse of the amount of time required to prefetch the document (the prefetch time). Therefore, *the flux product is inversely proportional to the prefetch time*. Accordingly, Internet resources with equal probabilities of access and component values will be prefetched in an order of ascending prefetch times. Independent claims 1, 25, and 27-29, as amended, require prefetching Internet resources according to a descending order of round trip times.

Thus, Horvitz does not disclose or suggest prefetching Internet resources according to a descending order of prefetch times, as required by independent claims 1, 17, 25, and 27-29, as amended. In fact, if anything, Horvitz actually teaches away from the present invention.

Kunkel et al. in view of Vaid et al. and Bryant et al.

The Examiner also asserts that Kunkel discloses the prefetching of Internet resources. The Examiner recognizes, however, the Kunkel does not disclose or suggest the prefetching of data based on a descending order of round trip times and data size.

5 The Examiner further asserts that Vaid teaches a system to schedule downloading of data in order to provide optimized computer usage. In addition, the Examiner notes that Vaid teaches “estimating a bit rate over a round-trip time between the data source and the data receiver.” *citing* Abstract.

10 While Vaid may disclose the estimation of round-trip times of data exchanged between a sender and a receiver in a data network, Vaid estimates the round-trip times of *TCP/IP packets*. The present invention, on the other hand, estimates the round-trip times of *HTTP request/response events*.

15 HTTP traffic travels through software that runs atop TCP/IP stacks on client, server, and proxy computers. HTTP traffic therefore incurs latencies caused by such software. These latencies are generally not reflected in the TCP/IP round-trip times computed by Vaid. Moreover, HTTP traffic for the purpose of Web document prefetching, as in the context of the present invention, shows characteristics in terms of length and origin server access frequency that general TCP/IP traffic does not exhibit (the origin server is referred to as the “receiver” in Vaid). The present invention takes advantage of such HTTP traffic characteristics in its round-trip time estimation.

20 For example, the present invention considers the length of an HTTP response (e.g., via an HTTP HEAD request) and the actually measured round-trip times of previous HTTP requests to the same origin server when computing the estimated round-trip time of an HTTP request/response event. Thus, the round-trip time estimations of the present invention are dynamically adjusted, based on changing network and server conditions, because estimates are based on previously measured round-trip times of HTTP request/response events to the same origin server. The adaptation to changing actual round-trip times is effected by a linear weighing function that uses previous HTTP request/response events, if available, to the same origin server as data points. If there was no previous HTTP request/response event to the same origin server in the past, a baseline estimate is established through actually fetching a document from the origin server via an HTTP

GET request rather than estimating the round-trip time of the document. Vaid uses an unspecified baseline estimate in all cases.

Thus, Kunkel et al. or Vaid et al., alone or in combination, do not disclose or suggest prefetching one or more Internet resources based on a descending order of estimated round trip times that is "based on an interval of time between a sending of an HTTP request and a receipt of a response to said HTTP request," as required by each of the independent claims, as amended.

An example helps to illustrate the fundamental differences between Vaid and the present invention. Assume that a user, A, is communicating with two servers, B and C. The time interval between A sending a TCP packet to B and receiving back an acknowledgement is 100 msec.

The time interval between A sending a TCP packet to C and receiving back an acknowledgement is 200 msec. Further assume that identical Web servers are added to both B and C and that B has a slower processor than C. An HTTP request for a resource, R, processed by the Web server on machine B takes 400 msec, whereas processing an HTTP request for a resource, R', by the Web server on C takes 100 ms. It is assumed that the HTTP requests for R and R' and the HTTP responses to R and R' all fit in one TCP/IP packet each, a quite typical scenario.

Thus, the estimated round trip time for sending the HTTP request for resource, R, from machine A to B (100 msec), processing the request in B's HTTP server (400 msec), and sending the HTTP response back to A (100 msec) is a total of 600 msec. The estimated round trip time for sending the HTTP request for resource, R', from machine A to C (200 msec), having C's HTTP server process the request (100 msec) and sending the response back to A (200 msec) is a total of 500 msec.

Thus, based on the TCP round trip time computed by Vaid, the resource from server C (with the longer TCP time) would be prefetched first. Based on the estimated round trip time computed by the present invention, however, the resource from server B (with the longer estimated HTTP round trip time) would be prefetched first. Thus, applying the present invention to this example yields exactly the opposite result that would be achieved using the round trip time of Vaid. However, Applicant asserts that this inventive method results in a more responsive and enjoyable user experience with reduced latencies in obtaining requested Web resources.

Additional Cited References

Narayanaswami was also cited by the Examiner for its disclosure that present Web

pages are resolved periodically so as to maintain a list of currently active links based on one or more variables.

Narayanaswami is directed to the dynamic multiplexing of hyperlinks and bookmarks. Narayanaswami do not address the issue of prefetching based on prefetch times.

5 Thus, Narayanaswami does not disclose or suggest prefetching Internet resources according to a descending order of prefetch times, as required by independent claims 1, 17, 25, and 27-29, as amended.

Bryant et al. were also cited by the Examiner for its disclosure of measure response times as seen by an end user for requests submitted from a Web browser to a Web server.

10 Bryant is directed to "a method of logging information in a computer network having a Web client connectable to a Web server. In response to the HTTP request,(and as a result of receiving a response to that request), a response time associated with that first HTTP request is calculated. Thereafter, the response time calculated is passed from the Web client on a subsequent HTTP request to that Web server, where the information is logged for later use."

15 Citing Abstract.

Thus, Bryant et al. do not disclose or suggest prefetching Internet resources according to a descending order of prefetch times, as required by independent claims 1, 17, 25, and 27-29, as amended.

Dependent Claims 2-16, 18-24 and 26

20 Dependent Claims 2-16, 18-24 and 26 were rejected under 35 U.S.C. §103(a) as being unpatentable over Horvitz et al. in view of Bryant et al., under 35 U.S.C. §103(a) as being unpatentable over Kunkel et al. in view of Vaid et al. and Bryant et al. and under 35 U.S.C. §103(a) as being unpatentable over Kunkel et al. in view of Narayanaswami and Bryant et al. Claims 2-16, 18-24 and 26 are dependent on Claims 1, 17 or 25, respectively, and are therefore patentably
25 distinguished over Horvitz et al., Kunkel et al., Vaid et al., Narayanaswami, and Bryant et al. (alone or in any combination) because of their dependency from amended independent Claims 1, 17 or 25, for the reasons set forth above, as well as other elements these claims add in combination to their base claim.

30 In view of the foregoing, the invention, as claimed in Claims 1 through 29, cannot be said to be either taught or suggested by Horvitz et al., Kunkel et al., Vaid et al., Narayanaswami, and

Bryant et al. (alone or in any combination). Accordingly, applicant respectfully requests that the rejection of claims 1 through 29 under 35 U.S.C. § 103(a) be withdrawn.

All of the pending claims, i.e., claims 1 through 29, are in condition for allowance and such favorable action is earnestly solicited.

5 If any outstanding issues remain, or if the Examiner has any further suggestions for expediting allowance of this application, the Examiner is invited to contact the undersigned at the telephone number indicated below.

The Examiner's attention to this matter is appreciated.

Respectfully,



10 Date: June 30, 2003

Kevin M. Mason

Attorney for Applicant(s)

Reg. No. 36,597

15

Ryan, Mason & Lewis, LLP

1300 Post Road, Suite 205

Fairfield, CT 06824

(203) 255-6560

VERSION MARKED TO SHOW ALL CHANGES

IN THE CLAIMS:

Please amend the claims, as follows:

1. (Amended) A method of prefetching one or more Internet resources referenced in one or more Web pages, said method comprising the steps of:

obtaining [an] one or more estimated round trip times for said Internet resources, wherein said one or more estimated round trip times [is] are based on an interval of time between a sending of an HTTP request and a receipt of a response to said HTTP request; and

prefetching said Internet resources based on a descending order of said one or more estimated round trip times.

negative limitation

2. (Unchanged) The method according to claim 1, wherein two or more of said Internet resources are prefetched substantially in parallel.

3. (Amended) The method according to claim 1, wherein said step of prefetching said Internet resources based on said one or more estimated round trip times is performed only for Internet resources associated with origin servers that have been previously accessed and said method further comprising the step of prefetching all Internet resources associated with servers that have not been previously accessed.

4. (Amended) The method according to claim 1, wherein said one or more estimated round trip times for each Internet resource is based on average access time statistics for the corresponding origin server and the actual size of said Internet resource when said actual size is available.

5. (Amended) The method according to claim 4, wherein said one or more estimated round trip times for each Internet resource is based on average access time statistics for the corresponding origin server and the average size of Internet resources provided by said origin server

if said origin server does not indicate said actual size.

6. (Amended) The method according to claim 4, wherein said one or more estimated round trip times for each Internet resource is based on average access time statistics for the corresponding origin server and the average size of Internet resources provided by said origin server if the setup and wait time for accessing said origin server is not significantly less than the average round trip time for Internet resources obtained from said origin server.

7. (Amended) The method according to claim 1, wherein said one or more estimated round trip times [is] are based on at least one actual prior round trip time for said Internet resource.

8. (Unchanged) The method according to claim 1, wherein said step of prefetching said Internet resources does not begin until said one or more Web pages have been fetched.

9. (Unchanged) The method according to claim 1, wherein said step of prefetching said Internet resources continues until said Internet resources have been prefetched or until a user selects a new Web page.

10. (Unchanged) The method according to claim 1, further comprising the steps of storing said Internet resources in a cache and determining if any of said Internet resources are already stored in said cache before prefetching begins.

11. (Unchanged) The method according to claim 1, further comprising the step of applying a filter to said Internet resources to reduce the overhead on network, server or local resources due to prefetching.

12. (Unchanged) The method according to claim 11, wherein said filter discards all Internet resources that do not use the HTTP protocol for transmission.

13. (Unchanged) The method according to claim 11, wherein said filter discards all

Internet resources that corresponding to dynamically generated Web resources.

14. (Unchanged) The method according to claim 11, wherein said filter discards all Internet resources that correspond to resources whose size is more than a certain maximum size threshold.

15. (Unchanged) The method according to claim 11, wherein said filter discards all Internet resources that correspond to resources whose estimated round trip time is longer than a certain maximum time.

16. (Unamended) The method according to claim 11, wherein said filter discards all Internet resources that correspond to resources whose estimated round trip time is shorter than a certain minimum time threshold.

17. (Amended) A method of prefetching one or more Internet resources referenced in one or more Web pages, said method comprising the steps of:

determining [an] one or more estimated round trip times for said Internet resources based on an interval of time between a sending of an HTTP request and a receipt of a response to said HTTP request;

sorting a list of said Internet resources based on a descending order of said one or more estimated round trip times;

prefetching said sorted list of Internet resources until one or more predefined threshold conditions are met.

18. (Unchanged) The method according to claim 17, wherein two or more of said Internet resources are prefetched substantially in parallel.

19. (Amended) The method according to claim 17, wherein said step of prefetching said Internet resources based on said descending order of said one or more estimated round trip times is performed only for resources associated with origin servers that have been previously accessed

and said method further comprising the step of prefetching all resources associated with servers that have not been previously accessed.

5 20. (Amended) The method according to claim 17, wherein said one or more estimated round trip times for each Internet resource is based on average access time statistics for the corresponding origin server and the actual size of said Internet resource when said actual size is available.

10 21. (Amended) The method according to claim 20, wherein said one or more estimated round trip times for each Internet resource is based on average access time statistics for the corresponding origin server and the average size of Internet resources provided by said origin server if said origin server does not indicate said actual size.

15 22. (Amended) The method according to claim 20, wherein said one or more estimated round trip times for each Internet resource is based on average access time statistics for the corresponding origin server and the average size of Internet resources provided by said origin server if the setup and wait time for accessing said origin server is not significantly less than the average round trip time for Internet resources obtained from said origin server.

20 23. (Unchanged) The method according to claim 20, further comprising the step of applying a filter to said Internet resources to reduce the overhead on network, server or local resources due to prefetching.

25 24. (Unchanged) The method according to claim 23, wherein said filter discards all Internet resources selected from the set comprised substantially of those Internet resources that (i) do not use the HTTP protocol for transmission; (ii) correspond to dynamically generated Web resources; (iii) correspond to resources whose size is more than a certain maximum size threshold, (iv) correspond to resources whose estimated round trip time is longer than a certain maximum time,
30 or (v) correspond to resources whose estimated round trip time is shorter than a certain minimum

time threshold.

25. (Amended) A system for prefetching one or more Internet resources referenced in one or more Web pages, each of said Internet resources having an associated origin server, said tool
5 comprising:

a memory for storing a server statistics database that records access time statistics for each origin server that has been previously accessed;

a processor operatively coupled to said memory, said processor configured to:

10 obtain [an] one or more estimated round trip times for said Internet resources, wherein said one or more estimated round trip times [is] are based on an interval of time between a sending of an HTTP request and a receipt of a response to said HTTP request; and

prefetch said Internet resources based on a descending order of said one or more estimated round trip times.

15 26. (Unchanged) The system according to claim 25, wherein said server statistics database records the average setup, wait and byte transmission times and average resource size for said Internet resources obtained from said corresponding origin server.

20 27. (Amended) A method of prefetching one or more Internet resources referenced in one or more Web pages, said method comprising the steps of:

determining if one or more of said Internet resources are candidates for prefetching based on [an] one or more estimated round trip times, wherein said one or more estimated round trip times [is] are based on an interval of time between a sending of an HTTP request and a receipt of a response to said HTTP request; and

25 prefetching said Internet resources that are determined to be candidates for prefetching according to a descending order of said one or more estimated round trip times.

28. (Amended) An article of manufacture for prefetching one or more Internet resources referenced in one or more Web pages, said article of manufacture comprising:

a computer readable medium having computer readable program code means embodied thereon, said computer readable program code means comprising program code means for causing a computer to:

5 obtain [an] one or more estimated round trip times for said Internet resources, wherein said one or more estimated round trip times [is] are based on an interval of time between a sending of an HTTP request and a receipt of a response to said HTTP request; and

prefetch said Internet resources based on a descending order of said one or more estimated round trip times.

10 29. (Amended) A method of prefetching one or more Internet resources referenced in one or more Web pages, said method comprising the steps of:

obtaining [an] one or more estimated round trip times for said Internet resources, wherein said one or more estimated round trip times [is] are based on an interval of time between a sending of an HTTP request and a receipt of a response to said HTTP request;

15 identifying a subset of said Internet resources that are candidates for prefetching based on said one or more estimated round trip times; and

determining whether to prefetch one or more of said Internet resources in said subset of Internet resources based on predefined conditions, at least one of said predefined conditions being based on a descending order said one or more estimated round trip times.

20